DOCUMENT RESUME

ED 259 013 TM 850 382

AUTHOR	McCormick, Bruce K.; And Others
TITLE	The Effects of Practice on the Armed Services Vocational Aptitude Battery.
INSTITUTION	Canyon Research Group, Inc., Westlake Village, Calif.
SPONS AGENCY	Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
REPORT NO	ARI-TR-602
PUB DATE	Dec 83
GRANT	MDA-903-82-M-3943
NOTE	32p.; Portions of this document were presented at the meeting of the Military Testing Association (San Antonio, TX, November 1982).
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Aptitude Tests; Drills (Practice); Individual Differences; *Military Personnel; Performance Factors; Personnel Selection; *Test Reliability; Test Results; Test Use; Test Validity; *Test Wiseness; *Vocational Aptitude
IDENTIFIERS	Armed Forces Qualification Test; *Armed Services Vocational Aptitude Battery

ABSTRACT

Alternate forms of the Armed Services Vocational Aptitude Battery (ASVAB) were administered five separate times to 57 men and women of military service age. The objective was to determine to what extent means and cross-session correlations are stable over several administrations. Ten individual subtests, the derived ASVAB area composites and the Armed Forces Qualification Test (AFQT) were examined for stability. The means for this sample were below the national average and scores were less dispersed. Means increased over sessions .5 standard deviation or more on half the subtests and consequently, on most of the composite scores. Correlations for the subtests and the composites were largely stable over sessions and were slightly higher later in practice. Reliabilities were comparable to reference populations when adjusted for the range restriction of the present sample. The implications of practice effects for paper and pencil as well as automated selection tests are discussed. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The Eric Facility has assigned this document for proccesing to: TM
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should refect their special points of view. CE

ED259013

Technical Report 602

The Effects of Practice on the Armed Services Vocational Aptitude Battery

Bruce K. McCormick and William P. Dunlap
Tulane University

Robert S. Kennedy
Essex Corporation

Marshall B. Jones
Hershey Medical Center

Selection and Classification Technical Area
Manpower and Personnel Research Laboratory

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

U. S. Army

Research Institute for the Behavioral and Social Sciences

December 1983

Approved for public release; distribution unlimited.

TH 850 382

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Research accomplished under contract
for the Department of the Army

Canyon Research Group, Inc.

Technical review by

M.A. Fischl
Rebecca L. Oxford-Carpenter
Stephen Cormier

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POT, 5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report 602	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): The Effects of Practice on the Armed Services Vocational Aptitude Battery	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report 18 Mar 1982 - 18 Mar 1983
	6. PERFORMING ORG. REPORT NUMBER: ARI-1
7. AUTHOR(s): McCormick, B.K., Dunlap, W.P. (Tulane Univ.); Kennedy, R.S. (Essex Corp.); Jones, M.B. (Hersey Medical Center)	8. CONTRACT OR GRANT NUMBER(s): MDA 903-82-M-3943
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Canyon Research Group, Inc. 741 Lakefield Road, Suite B Westlake Village, CA 91361	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q 162722A791
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Research Institute for the Behavioral and Social Science Alexandria, VA 22333-5600	12. REPORT DATE: December 1983 / 13. NUMBER OF PAGES: 31
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified / 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Portions were presented at the meeting of the Military Testing Association, San Antonio, November 1982.

Technical quality of this research was monitored by Hilda Wing.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ASVAB	Classification	Individual Differences
Selection	Validity	Tests and Measurements
Practice Effects	Performance Stability	Cognitive Tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Alternate forms of the ASVAB were administered five separate times to fifty-seven men and women of military service age. The objective was to determine to what extent means and cross-session correlations are stable over several administrations. Ten individual subtests, the derived ASVAB area composites and the Armed Forces Qualification Test (AFQT) were examined for

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

stability. The means for this sample were below the national average and scores were less dispersed. Means increased over sessions .5 standard deviation or more on half the subtests and consequently, on most of the composite scores. Correlations for the subtests and the composites were largely stable over sessions and were slightly higher later in practice. Reliabilities were comparable to reference populations when adjusted for the range restriction of the present sample. The implications of practice effects for paper and pencil as well as automated selection tests are discussed.

Technical Report 602

The Effects of Practice on the Armed Services Vocational Aptitude Battery

Bruce K. McCormick and William P. Dunlap
Tulane University

Robert S. Kennedy
Essex Corporation

Marshall B. Jones
Hershey Medical Center

Hilda Wing, Contracting Officer's Representative

Submitted by
Newell K. Eaton, Chief
Selection and Classification Technical Area

Approved technically adequate
and submitted for publication by
Joyce L. Shields, Director
Manpower and Personnel
Research Laboratory

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

December 1983

Army Project Number
2Q162722A791

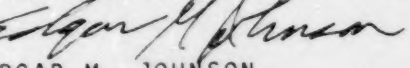
Manpower and Personnel Research

Approved for public release; distribution unlimited.

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The Armed Services Vocational Aptitude Battery (ASVAB) is a multiaptitude test battery used for selection and classification of United States Military personnel. The purpose of this research was to examine the effects of massed retesting or practice on the statistical characteristics of ASVAB subtest and composite scores. Applicants who fail to qualify because of low ASVAB scores may be permitted to retake the test battery. The results of this research showed the level (means) of test scores to increase somewhat over sessions but other characteristics of the battery (variances, reliabilities, covariances) remained stable, after correction for range restriction. That is, individuals will probably improve their scores with retesting, but the psychometric properties of those improved scores are not changed.



EDGAR M. JOHNSON
Technical Director

THE EFFECTS OF PRACTICE ON THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY

EXECUTIVE SUMMARY

Requirement:

To study the stability of the statistics of the Armed Services Vocational Aptitude Battery (ASVAB) over multiple administrations.

Procedure:

Five alternate forms of the ASVAB were administered to fifty-seven men and women of military service age. The objective was to determine to what extent means and cross-session correlations are stable over several administrations. Ten individual subtests and combinations of certain of these subtests were examined for stability.

Findings:

The means for this sample were below the national average, and scores were less dispersed. Means increased over sessions .5 standard deviation or more on half the subtests and, consequently, on most of the composite scores. Correlations for the subtests and the composites were largely stable over sessions and were slightly higher later in practice. Reliabilities were comparable to reference populations when adjusted for the range restriction of the present sample. The implications of practice effects for paper and pencil, as well as automated, selection tests are discussed.

Utilization of Findings:

These analyses provide evidence for the differential stability of composites formed from the ASVAB. The trend toward increasing means with extended practice should be replicated in a larger, more representative sample. If cross-validated, such a replication will recommend the requirement for accurate record keeping of prior ASVAB testing of applicants for military service.

THE EFFECTS OF PRACTICE ON THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY

CONTENTS

Page

INTRODUCTION..... 1

METHOD..... 4

RESULTS..... 6

DISCUSSION..... 9

REFERENCES..... 13

APPENDIX A..... 22

A Comparison of Test/Retest Correlations for the Present Sample (N=57) Early (Sessions 1, 2) and Late (Sessions 4, 5) in Practice and a Reference Sample Tested Twice

TABLES

Table 1. Aptitude Area Composites Used in ASVAB 8/9/10.. 16

2. Means and Standard Deviations for Five Successive Test Administrations Ordered by Strength of Linear Trend with Linear and Quadratic Probabilities for 10 Subtests..... 17

3. Cross-Session Correlations of the Ten Test Scores (N=57)..... 18

4. Means and Standard Deviations for Five Successive Test Administrations Ordered by Strength of Linear Trend with Linear and Quadratic Probabilities for 10 Subtests..... 19

5. Inter-Administration Correlations of the Ten Area Composites and AFQT..... 20

THE EFFECTS OF PRACTICE ON THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY

INTRODUCTION

Several years ago, Jones (1969) proposed a two process theory to describe individual differences in the acquisition of skills. No inference was made at that time concerning the potential relevance of that theory to changes in tests of ability. The theory posited an acquisition phase, in which persons improved at different rates, and a terminal phase, in which persons reach or approximate their individual limits. The theory specified that different persons could be expected to begin at different points initially, and arrive at their different terminal levels via different pathways. The theory further implied that, after the terminal process is reached, persons will cease to change positions relative to each other, despite additional practice. In other words, several individuals may approach a task with differing experience levels and capacities, both of which influence their initial scores. As practice continues, previous experience will begin to contribute proportionately less to a person's score, and individual differences in learning would increasingly influence his/her test score. As the amount of experimental time increases proportional to previous practice, and as learning progresses, differences between subjects will become more attributable to actual differences in underlying capacity or ability, until finally, the amount of ability is largely what governs performance scores.

Thus, an inter-session correlation matrix would present a distinctively different appearance if performance early versus late in practice were examined. Early in practice, one would observe the superdiagonal form (Jones, 1969), in which correlations between adjacent trials would be higher than comparisons which are more remote. If the theory holds, the cross-session correlation coefficients would eventually become constant and symmetrical. When this occurs, no systematic differences would be present in the matrix as a function of temporal separation. If the terminal process is not reached, then the matrix will continue to show superdiagonal form (Jones, 1969), and the task is considered not to have stabilized.

Recently, a program was begun to standardize a performance test battery applying these principles of differential stability (Kennedy & Bittner, 1977). In order to study the effects to humans of adverse environments, it would be desirable that the test battery assess complex mental abilities which could be related to elements of military jobs. A natural consequence of research in this area of environmental stress is that, generally, each subject serves as his own control over many sessions. In other words, repeated measures analysis of variance is required--a differential approach. Moreover, within the context of this

theory, performance on all tasks in the battery would need to be at terminal levels before an experimental treatment was introduced. Many batteries have purported to measure primary mental abilities, and several have been factor analyzed. However, none of these had been examined in terms of stability of subtests over sessions; and, generally, the factor analyses which were performed were conducted on, at most, two replications, a questionable approach if rate changes occur due to practice (cf., Alvares & Hulin, 1972).

Findings from over sixty tests (Kennedy, Carter & Bittner, 1980), which were administered in a fifteen-day repeated measures paradigm, support the rate-terminal theory of skill acquisition. Additionally, these findings permit the theory to be generalized to include other behavioral tests. Specifically, the data indicate that people do exhibit differential rate processes over practice, when faculties are measured by tests of short-term memory, grammatical reasoning, learning ability, and several other cognitive tests (see Kennedy & Harbeson, 1981, for a review).

Researchers have studied practice effects on intelligence and ability tests, and it has been known since at least 1920 that test scores increase (Dunlap & Snyder, 1920; Gundlach, 1926; Thorndike, 1922). Additionally, reviews of performance changes on individually administered intelligence tests (Thompson, 1975) and scholastic aptitude (Nader, 1980), when administered over repeated testings, have suggested that performance on these tasks also may be less stable than previously considered.

In recent years, there has been an increased interest in practice and coaching effects (Anastasi, 1981; Catron & Thompson, 1979; Messick & Jungblut, 1981; Whimbey, Carmichael, Jones, Hunter & Vincent, 1980; Wing, 1980). However, few investigations have been conducted which involve more than two or three replications. What evidence there is suggests that repeated testing may produce appreciable effects on mean test scores. Mackaman, Bittner, Harbeson, Kennedy and Stone (1982) found that inter-session correlations on the Wonderlic were stable over 18 replications, but the scores increased, on the average, 21 percentile points. This suggests that exposure history is an important variable, with regards to the testing and subsequent assignment of personnel.

The Armed Services Vocational Aptitude Battery (ASVAB) possesses many of the same type of test items as the Wonderlic (Kass, Mitchell, Grafton & Wing, 1982). In addition, other tests, similar to the subtests found in ASVAB, have not always differentially stabilized after many trials (cf., Kennedy et al., 1981), and rarely have tests exhibited mean or differential stability from the first session. The importance of this lack of stabilization should not be overlooked. Various combinations of ASVAB subtests are used for counseling (Fischl, Foss & McBride, 1979) and for assignment to service schools (Sims & Hiatt, 1981;

Swanson, 1979). In a review of 95 different Navy enlisted ratings, Carter and Biersner (1982) showed how abilities from ASVAB and other aptitude test batteries would map onto disparate Navy jobs. If a test were unstable, then predictions made on the basis of scores from it would be less accurate. Thus the value of prediction would be lessened.

Various subgroups of the population with whom the ASVAB is used may vary with respect to amount of experience in taking standardized tests. It might be expected that individuals with less sophistication in test-taking skills would take longer to produce a stable pattern of scores. Moreover, the initial test scores of these individuals would be less effective in predicting later performance. Additionally, racial differences in repeated measures of test performance were reported by Dyer (1970). He found that in uncoached practice sessions, black college students showed a statistically significant increase over white students in three administrations of alternate forms of a standardized test of reasoning ability. An investigation of repeated administrations of the ASVAB, therefore, should include examination of performance which may be unique to particular groups of individuals with whom the test may be used.

It was the purpose of this investigation to determine whether practice modified performance on alternate forms of the ASVAB. Practice effects would be observed as changes in means, variances and cross-session correlations. Stability of ASVAB would be determined according to the extent to which the test met standards developed in repeated measures experimentation and included group and differential criteria. It was hypothesized that improvement would continue over sessions, and that some tests would be differentially unstable.

METHOD

Subjects

The subjects were 57 men and women enrolled as trainees in the Job Corps Center, Shreveport, LA. Thirty-four subjects were male (29 Black and 5 White), and 23 were female (19 Black and 4 White). Effort was made to assure maximum response by Center trainees. It was explained that subjects would be required to take the ASVAB on five consecutive mornings and that the results would be used for research purposes. Additionally, trainees were told that their scores from the first day of testing could be used for determining their eligibility for enlistment in the armed services, if they so desired. It was emphasized that participation in this project would not obligate subjects to consideration for military service. Trainees were also told that they would be paid for their participation contingent upon completion of all five days of testing. The first 60 volunteers were selected. On the second day of testing, two subjects dropped out, and a third quit on the fourth day. All three left due to unforeseen work, school or family circumstances.

Apparatus and Procedure

Five forms of the ASVAB were administered from 8:00 AM to 12:00 noon in a group setting for five consecutive days. On each day of testing all subjects took the same form of the ASVAB. The order of administration was: Form 8b, 9a, 9b, 10a, 10b. These five forms are considered of equal difficulty (Ree, Mullins, Mathews & Massey, 1982). Forms of the ASVAB having the same number also had identical items comprising the subtests of:

General Science (GS)	Mathematics Knowledge (MK)
Coding Speed (CS)	Mechanical Comprehension (MC)
Auto/Shop Information (AS)	Electronics Information (EI)

Different across forms were:

Paragraph Comprehension (PC)	Numerical Operations (NO)
Arithmetic Reasoning (AR)	Word Knowledge (WK)

For additional information the reader is referred to the reference works of Ree et al. (1982) and Kass et al. (1982). Administration followed standard procedures and was conducted by members of the Shreveport Military Enlistment Processing Station (MEPS). Neither coaching nor feedback was given to subjects during the days of testing.

Scoring

Subjects' responses were made on answer sheets which were scored by computer at the MEPS or the afternoon of each day of the project. ASVAB subtest results were reported in raw score form. These different subtests were combined to form composite scores for AFQT and for ten aptitude areas. (See Table 1.)

RESULTS

ASVAB Subtests

Means

Significant linear trend, indicating an improvement with practice in the absence of feedback, occurred with four test sections: Coding Speed, Numerical Operations, Mathematics Knowledge and Mechanical Comprehension.[1] The means and associated p-values for linear and quadratic relationships are presented in Table 2. The most dramatic increases were for Coding Speed and Numerical Operations, where the average fifth test performance exceeded the average of the first test performance by 48.3% and 27.0%, respectively. No test showed a significant drop with practice. However, both Word Knowledge and Paragraph Comprehension showed significant quadratic (U-shaped) changes over sessions, which suggests possible motivational deficits on the intermediate Days 2, 3 and 4. The significant quadratic component for Coding Speed was apparently due to the rapid increase in mean score from Day 1 to Day 2, followed by a slower increase thereafter.

The mean scores on the first administration are slightly more than one standard deviation below those reported by others (Kass et al., 1982; Ree et al., 1982). However, of those tests which later showed improvement (viz., CS, NO, MK, MC), the arithmetic mean scores are slightly less than a standard deviation lower in subsequent sessions than found in these other experiments. The standard deviations were constant over sessions and about 75% the size of the larger samples (Kass et al., 1982; Ree et al., 1982).

Correlations

The intercorrelations across five repeated administrations of each subtest of the ASVAB are presented in Table 3. The sample size obtained (N = 57) was too small to permit reliable inferences from factor analyses.

For five of the tests (General Science, Arithmetic Reasoning, Word Knowledge, Numerical Operations and Coding Speed), the highest correlations approximate conventional reliability estimates. However, for the remaining five tests (Paragraph Comprehension, Auto/Shop Information, Mathematics Knowledge, Mechanical Comprehension and Electronics Information), the "highest" figures are lower than conventional reliability estimates (cf., Kennedy et al., 1980). The latter five tests are stable in the sense that all five administrations measure the same underlying variation (cf., Jones & Kennedy, 1983). The ASVAB composites, as would be expected, have much higher reliabilities and intersession correlations (see Table 4).

---

[1] One or more of these subtests are included in nine of the ten composites--the exception being GT.

The correlations improved over the five practice sessions for nine out of ten subtests, the exception being Electronics Information. The average intersession correlation for the first three days (1, 2 and 3) was compared to the average of the last three days (i.e., 3, 4 and 5). It is recognized that, while not an independent comparison, it is instructive to compare the means. The mean improvement in reliability correlation was small (viz., r=.61 versus .68) but obvious, and in some cases non-trivial (e.g., CO r =.72 vs .84). The correlations, corrected for attenuation due to range restriction following the equation in Sims and Hiatt (1981), are consistent with those reported in Friedman, Streicher, Wing and Grafton (1982). The later days' correlations (days 4 and 5) are all slightly higher, the early days' (1 and 2) approximately the same or else higher. These values appear in Appendix A.

Sex

Sex only approached significance on one subtest, Mechanical Comprehension, $F(1,53) = 3.25$, $p = .0772$; the mean for females was 7.71 and the mean for males was 9.34.

ASVAB Composites

Means

Linear and quadratic trends are reported in Table 4. Significant trends occurred for the Armed Forces Qualification Test (AFQT) score and for all composites but General Technical (GT) and Skilled Technical (ST). In the case of General Maintenance (GM) and Electronics Repair (EL), the increase was small but significant (< .2 standard deviation). In the first session, the composite score which occurred one standard deviation above this group's mean was 76.2. After five sessions, the composite score one standard deviation above this group's mean was 80.6 ($p<.001$).

Correlations

Table 4 contains the cross-session correlations for the ten area composites and for AFQT. The overall impression is of high correlations and general stability, although the average intersession correlation for the last three days (3, 4 and 5) is, in all cases but one (Surveillance/Communications), higher than the average intersession correlation for the first three days (i.e., 1, 2 and 3).

Summary of Results

The means and dispersions of scores for this population were below the national average. On half the subtests, means increased over sessions .5 standard deviation or more and, consequently, on most of the composite scores. Correlations for the subtests and the composites were largely stable over sessions and were slightly higher later in practice. Reliability correlations were comparable to reference populations when adjusted for the range restriction of the present sample.

DISCUSSION

Stability

The original purpose of the present research was to determine whether repeated administrations of forms of the ASVAB would produce evidence of stability of scores. This question is of interest for selection, classification and prediction in general, but these issues have different relevance, depending on whether representative or exceptional populations are studied. The availability of a small (N=60) Job Corps group encouraged us to research this question in such a population. It was recognized that information derived from a homogeneous sample would be less generalizable than one which would be more heterogeneous. However, the increase in mean scores which was expected to occur in such a group might be more likely to emphasize transition across boundaries of administrative decisions (e.g., selection cut-off scores and service school assignment).

ASVAB was administered five separate times to fifty-seven men and women of military service age. Ten individual subtests, the derived ASVAB area composites (N=10) and the Armed Forces Qualification Test (AFQT) were examined for group and differential stability. The means and dispersions of scores for this sample were below the national average. Means increased over sessions .5 standard deviation or more on half the subtests and, consequently, on most of the composite scores. In the present experiment, differential stabilization (Jones, Kennedy & Bittner, 1981) with practice does not appear to be a problem in ASVAB. All ten subtests were more or less differentially stable on the first administration. The same was true for the ten aptitude area composites. In neither the subtests nor the area composites was there any appreciable differential change with practice, although mean changes on repeated administrations of the ASVAB did occur.

Mean changes are an index of group stability. Four of the subtests showed significant increasing linear trend with practice. Four of the area composites showed increases from the first to the fifth administration of .5 standard deviation or more. These changes are sufficient to warrant some concern, although they are not surprising in light of the Mackaman et al. (1982) finding of almost 21 percentile points improvement with practice in a population whose mean score began at the 50th percentile. For example, if II were used as a cutoff score for AFQT: a) 1/3 of those in the present experiment who initially failed to achieve this score later surpass this score at least once; and b) 1/6 of them would pass more than once.

Two questions about time lapse need to be answered: Whether the same sort of improvement would occur: a) if the five administrations of the present study were distributed over weeks

or months, instead of days? and b) in a more representative sample? We would predict that the present improvement is near optimum,or might be better if administered within one month. In our view, similar relative improvements (standard scores) would be observed in a more representative population.

A word, perhaps, is in order regarding the possibility that the results observed are due to regression. Men and women who enter the Job Corps do so, at least in part, because of poor performance in school and on the job. They are selected, if you like, on the basis of previous poor performance. To the extent that this previous poor performance may have involved transient (error) components, the possibility exists that the average error score in the sample studied may be negative at first testing. If so, the group mean would be expected to increase at retesting, as observed. However, it would not be expected to increase regularly with subsequent testing, as also happened. The possibility of a regression effect cannot, therefore, be excluded; but it seems unlikely to account for more than part of the observed increase with multiple retesting.

Implications for Selection

The Armed Forces Qualification Test (AFQT) score is employed in preliminary screening. It is used to classify individuals into five mental categories in order to determine eligibility for enlistment and particular job training (Mathews & Ree, 1982). Sims and Hiatt (1981) concluded that 83% of the predictive efficiency of the ASVAB is contained within the AFQT. Were abbreviated versions of the ASVAB created in order to screen individuals for more comprehensive testing, it is likely that these subtests, or ones like them, would be candidates for automated test administration through microcomputer. Therefore, it may be advisable to determine whether such improvement would occur on AFQT scores in a sample whose mean scores are more nearly like those for average Army recruits.

It should be noted, however, that this improvement should not be considered to be evidence of differential instability. If the latter were to occur, persons who scored lower initially might score higher later, and the converse. In the present experiment, the movement of subjects toward increasing scores with practice was largely uniform. Therefore, if movement across boundaries is a problem for ASVAB utilization, it will be necessary to monitor the number of times the test is taken. Thus, better predictive validities might be available from later test performances, because the correlations are higher.

Suggestions for Future Research

Several of the correlations for aptitude area composites tend to increase with practice, a finding which has been reported many

times before in repeated measures testing (cf., Kennedy et al., 1981). We do not believe that the restricted range of the present sample influenced this improvement; however, this finding should be checked. The result implies that improved reliability correlations might be available in later sessions; such improvement may be useful for classification. It is possible that certain persons may profit more than others by extra test taking. For example, persons new to test taking, who may qualify as borderline acceptable for the military service schools which have less stringent requirements, could be misassigned to these latter occupations when they could also be successful in more demanding jobs. While it is recognized (Schmidt & Hunter, 1981) that "selecting from the top down maximizes the productivity of employees selected" (p. 1130), those same authors propose greater relevance for a classification model than a selection model (Hunter & Schmidt, 1982). According to this view, individuals should be assigned to jobs based on the criterion of maximizing productivity. The prospect that improved differential predictive validities from disparate composites may be available with increased practice on the ASVAB subtests suggests that such an investigation should be performed with a larger sample than we used, and should include persons who are more representative of an incoming military population and with longer time intervals. It is not unlikely that extra testing might expand the service pool (Sims & Hiatt, 1981) from the standpoint of successful service school assignment.

Future Trends in Testing

Although paper and pencil tests of cognitive ability have strong roles in selection and classification, the advent of microprocessors likely will have an influence on automating future efforts in this area. If test automation of ASVAB proceeds further than simply translating the existing tests to microcomputer/video format, it may be helpful to study practice effects. This helpfulness depends on exploiting the possibilities of the new technology by developing new tests, tests that involve more elements of a perceptual, information processing, psychomotor and decision-making sort. Indeed, it is considered in some places (e.g., O'Leary, 1979) that a "job sample" approach not only has a higher likelihood of success, but is more apt to be fair than some of the tests which are now employed in selection. In view of the difficulties in the use of paper and pencil tests in classification (cf., Eaton, Bessemer & Kristiansen, 1979), it is suggested that video games have strong prospects to fill such a role. In one experiment (Lintern & Kennedy, 1982), which was later cross-validated (Westra, 1983), a video game correlated with a full-scale simulation of a night carrier landing as much as the test-retest reliability of the criterion would allow. It is offered that microcomputer video games might provide a fertile target of opportunity (Jones, Kennedy & Bittner, 1981). It should be noted that, when automated, these and other such tests usually involve implicit

knowledge of results, which might be expected to show greater changes in the mean than were found in the present research. Consequently, it is likely that, with practice, they will show appreciable differential change (Jones, 1981), as well. The promising possibility of introducing more heterogeneity into the ASVAB also will probably revive stabilization-with-practice as a major concern.

REFERENCES

Alvares, K. M. & Hulin, C. L. Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. Human Factors, 1972, 14, 295-308.

Anastasi, A. Coaching, test sophistication, and developed abilities. American Psychologist, 1981, 36, 1086-1093.

Carter, R. C. & Biersner, R. J. U. S. Navy enlisted jobs: An analysis (NBDL 82R010). New Orleans, LA: Naval Biodynamics Laboratory, May 1982.

Catron, D. W. & Thompson, C. C. Test-retest gains in WAIS scores after four retest intervals. Journal of Clinical Psychology, 1979, 35, 352-357.

Dyer, P. J. Effects of test conditions on negro-white differences in test scores. Doctoral thesis, Columbia University, New York, 1970.

Dunlap, K. & Snyder, A. Practice effects in intelligence tests. Jurnal of Experimental Psychology, 1920, 3, 396-403.

Eaton, N. K., Bessemer, D. W. & Kristiansen, A. M. Tank crew position assignment (Technical Report No. 391). Fort Knox, KY: Army Research Institute, 1979.

Fischl, M. A., Ross, R. M. & McBride, J. R. Development of factorially based ASVAB high school composites (Technical Paper No. 360). Alexandria, VA: Army Research Institute, April 1979.

Friedman, D., Streicher, A., Wing, H. & Grafton, F. Assessment of practice effects: Test-retest scores for FY81 Army applicants on ASVAB 8/9/10. Presented at the Military Tesing Association Annual Meetings, San Antonio, TX, September 1982.

Gulliksen, J. Theory of mental tests. New York: Wiley, 1950, p. 111.

Gundlach, R. The effects of practice on the correlations of three mental tests. The Journal of Educational Psychology, 1926, 127, 387-400.

Hunter, J. E. & Schmidt, F. L. Fitting people to jobs: The impact of personnel selection on national productivity. In M. D. Dunnette & E. A. Fleishman (Eds.). Human performance and productivity: Human capability assessment. Hillsdale, NJ: Lawrence Erlbaum Assoc., 1982, 233-284.

Jones, M. B. Differential processes in acquisition. In E. A. Bilodeau & I. McD. Bilodeau (Eds.). Principles of skill acquisition. New York: Academic Press, 1969, 141-170.

Jones, M. B., Kennedy, R. S. & Bittner, A. C., Jr. A video game for performance testing. American Journal of Psychology, 1981, 94, 143-152.

Jones, M. B. Further studies on stabilization with practice in performance tests. Final report (Contract No. N00203-80-M-5397). New Orleans, LA: Naval Biodynamics Laboratory, 15 May 1981.

Jones, M. B. & Kennedy, R. S. Stabilization with practice: Theory and assessment. Submitted to British Journal of Mathematical and Statistical Psychology, January 1983.

Kass, R. A., Mitchell, K., Grafton, F. & Wing, H. Factor structure of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9 & 10: Army applicant sample (Technical Report No. 440). Alexandria, VA: Army Research Institute, December 1982.

Kennedy, R. S. & Bittner, A. C., Jr. The development of a performance evaluation test for environmental research (PETER). In L. T. Pope & D. Meister (Eds.). Productivity Enhancement: Personnel Performance Assessment in Navy Systems (Technical Report No. AD A045047). San Diego, CA: Naval Personnel Research and Development Center, October 1977.

Kennedy, R. S., Carter, R. C. & Bittner, A. C., Jr. A catalogue of Performance Evaluation Tests for Envronmental Research. Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, CA, 13-17 October 1980, 344-348.

Kennedy, R. S. & Harbeson, M. M. A retrospective view of the "PETER" program. In J. C. Guignard & M. M. Harbeson (Eds.). Proceedings of the International Workshop on Research Methods in Human Motion and Vibration Studies. New Orleans: Naval Biodynamics Laboraory, September 1981.

Kennedy, R. S., Bittner, A. C., Jr., Carter, R. C., Krause, M., Harbeson, M. M., McCafferty, D. B., Pepper, R. L. & Wiker, S. F. Performance Evaluation Tests for Environmental Research (PETER): Collected papers (Report No. 80R008). New Orleans, LA: Naval Biodynamics Laboratory, July 1981.

Lintern, G. & Kennedy, R.S. A video game as a covariate for carrier landing research. Presented at the Eighth Psychology in the DOD Symposium, USAF Academy, Colorado Springs, CO, 21-23 April 1982.

Mackaman, S. L., Bittner, A. C., Jr., Harbeson, M. M., Kennedy, R. S. & Stone, D. A. Performance Evaluation Tests for Environmental Research (PETER): Wonderlic Personnel Test. Psychological Reports, 1982.

Mathews, J. J. & Ree, M. J. Enlistment Screening Test - Forms 81a and 81b: Development and calibration (Human Resources Technical Report No. 81-54). Brooks AFB, TX, March 1982.

Messick, S. & Jungblut, A. Time and method in coaching for the SAT. Psychological Bulletin, 1981, 89, 191-216.

Nader releases ETS report, hits tests as poor predictors of performance. Monitor, American Psychological Association, February, 1980, pp. 1,7.

O'Leary, L. A. Fair employment, sound psychometric practice, and reality. American Psychologist, 1973, 28, 147-150.

Ree, M. J., Mullins, C. J., Mathews, J. J. & Massey, R. H. Armed Forces Vocational Aptitude Battery: Item and factor analysis of Forms 8, 9 & 10 (AFHRL Report No. 81-55). Brooks AFB, TX: Air Force Human Resources Laboratory, March 1982.

Schmidt, F. L. & Hunter, J. E. Employment testing: Old theories and new research findings. American Psychologist, 1981, 36, 1128-1137.

Sims, W. H. & Hiatt, C. M. Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 & 7 with applications to ASVAB forms 8,9 & 10 (Report No. 1160). Alexandria, VA: Center for Naval Analyses, February, 1981.

Swanson, L. Armed Services Vocational Aptitude Battery Forms 6 & 7: Validation against school performance in Navy enlisted schools (July 1976-February 1978) (Technical Report No. 80-1). San Diego, CA: Navy Personnel Research and Development Center, Nov. 1979.

Thompson, C. C. The effects of practice on intelligence tests. Unpublished manuscript, Wake Forest University, 1973.

Thorndike, E. L. Practice effects in intelligence tests. Journal of Experimental Psychology, 1922, 5, 101-107.

Westra, D. P. Investigation of simulator design features for the carrier landing task (Technical Report No. 78-C-0060-7). Orlando, FL: Naval Training Equipment Center, 1982.

Whimbey, A., Carmichael, J. W., Jones, L. W., Hunter, J. T. & Vincent, H. A. Teaching critical reading and analytical reasoning in Project Soar. Journal of Reading, Oct. 1980, 6-9.

Wing, H. Practice effects with traditional mental test items. Applied Psychological Measurement, 1980, 4, 141-155.

Wing, H. Practice effects with traditional mental test items; A replication. Presented at the Symposium, "Practice Effects - Problems and Solutions," National Council on Measurement in Education Annual Meeting, Los Angeles, CA, April 1981.

Table 1
APTITUDE AREA COMPOSITES USED IN ASVAB 8/9/10

Aptitude Area Composite	Subtest Used in Computing Composites for ASVAB 8/9/10[1]
Combat (CO)	AR+AS+MC+CS[2]
Field Artillery (FA)	AR+MK+MC+CS
Electronics (EL)	AR+EI+MK+GS
Operators/Foods (OF)	NO+VE[3]+MC+AS
Surveillance/Communications (SC)	NO+CS+VE+AS
Motor Maintenance (MM)	NO+EI+MC+AS
General Maintenance (GM)	MK+EI+GS+AS
Clerical (CL)	NO+CS+VE
Skilled Technical (ST)	VE+MK+MC+GS
General Technical (GT)	VE+AR

Note: Table adapted from a table originally developed by Ms. Frances Grafton and Dr. Milt Maier.

1 Standard subtest scores are used in computation.

2 Abbreviations stand for the following:

AR	Arithmetic Reasoning
AS	Auto & Shop Information
CS	Coding Speed
EI	Electronics Information
MC	Mechanical Comprehension
MK	Math Knowledge
NO	Numerical Operation
GS	General Science

3 Verbal (VE) is a standard score conversion of the sum of raw scores for word knowledge (WK) and paragraph comprehension (PC).

Table 2
MEANS AND STANDARD DEVIATIONS FOR FIVE SUCCESSIVE TEST ADMINISTRATIONS ORDERED BY STRENGTH OF LINEAR TREND WITH LINEAR AND QUADRATIC PROBABILITIES FOR 10 SUBTESTS

Means

Section	8B	9A	9B	10A	10B	Linear	Quad
Coding Speed (CS)	26.7	33.6	36.0	34.9	39.6	.0000	.0081
Numerical Oper (NO)	24.1	26.4	26.4	29.4	30.6	.0000	.7602
Math Know (MK)	6.8	6.7	7.2	8.4	7.7	.0017	.7758
Mech Comp (MC)	8.1	7.7	7.6	8.7	8.5	.0200	.1316
Auto & Shop Info (AS)	7.3	7.6	7.8	7.8	8.1	.0757	.6083
Gen Science (GS)	8.6	8.1	7.9	8.0	8.1	.0824	.2023
Word Know (WK)	13.0	12.7	12.4	10.6	13.1	.1698	.0041
Electronics Info (EI)	6.5	6.3	6.7	6.8	7.0	.2280	.4727
Arithmetic Reas (AR)	8.9	8.4	10.0	9.4	9.3	.2730	.3533
Paragraph Comp (PC)	6.6	4.9	5.6	4.8	6.7	.5982	.0001

Standard Deviations

Section	8B	9A	9B	10A	10B
Coding Speed (CS)	13.9	15.2	16.6	14.7	15.2
Numerical Oper (NO)	9.4	11.1	9.2	10.9	10.9
Math Know (MK)	2.7	2.3	2.7	2.6	3.0
Mech Comp (MC)	3.2	3.0	3.2	3.3	3.6
Auto & Shop Info (AS)	2.7	3.4	3.0	3.3	3.2
Gen Science (GS)	3.5	3.8	3.9	3.8	4.0
Word Know (WK)	5.4	6.0	6.1	5.1	5.5
Electronics Info (EI)	2.5	2.9	2.4	3.2	3.0
Arithmetic Reas (AR)	3.9	3.3	3.4	3.5	4.2
Paragraph Comp (PC)	3.0	2.7	3.3	2.9	2.9

Table 3
CROSS-SESSION CORRELATIONS
OF THE TEN TEST SCORES N=57

General Science

	Session 2	3	4	5
Ses. 1	.66	.68	.72	.72
2		.74	.73	.70
3			.76	.79
4				.82

Arithmetic Reasoning

	Session 2	3	4	5
Ses. 1	.46	.56	.71	.63
2		.54	.51	.65
3			.58	.70
4				.73

Word Knowledge

	Session 2	3	4	5
Ses. 1	.70	.73	.67	.79
2		.80	.71	.80
3			.78	.83
4				.77

Paragraph Comprehension

	Session 2	3	4	5
Ses. 1	.62	.59	.58	.58
2		.69	.47	.69
3			.60	.66
4				.57

Numerical Operation

	Session 2	3	4	5
Ses. 1	.85	.86	.85	.86
2		.90	.90	.87
3			.90	.86
4				.93

Coding Speed

	Session 2	3	4	5
Ses. 1	.80	.73	.73	.67
2		.86	.82	.77
3			.85	.80
4				.86

Auto & Shop Information

	Session 2	3	4	5
Ses. 1	.44	.45	.66	.67
2		.58	.41	.39
3			.47	.54
4				.72

Mathematical Knowledge

	Session 2	3	4	5
Ses. 1	.24	.46	.52	.50
2		.39	.36	.26
3			.54	.42
4				.46

Mechanical Comprehension

	Session 2	3	4	5
Ses. 1	.20	.58	.38	.46
2		.48	.45	.40
3			.57	.56
4				.76

Electronics Information

	Session 2	3	4	5
Ses. 1	.64	.51	.54	.52
2		.54	.61	.54
3			.43	.35
4				.70

Table 4

MEANS AND STANDARD DEVIATIONS FOR FIVE SUCCESSIVE TEST ADMINISTRATIONS ORDERED BY STRENGTH OF LINEAR TREND WITH LINEAR AND QUADRATIC PROBABILITIES FOR 10 COMPOSITES

Means

	8b	9a	9b	10a	10b	Linear	Quad
AFQT	12.6	12.5	13.6	12.0	15.5	.0000	.0004
CL	69.3	72.5	72.8	73.4	79.0	.0000	.0752
MM	65.2	66.2	67.0	70.2	70.8	.0000	.6040
SC	66.5	68.9	70.3	70.0	74.5	.0000	.3168
CO	66.2	69.1	70.3	70.7	72.4	.0000	.2456
FA	68.6	72.2	73.4	75.6	76.1	.0000	.0890
OF	64.9	64.7	65.8	66.8	70.0	.0000	.0080
GM	64.9	65.1	65.1	67.0	66.9	.0112	.6220
EL	66.2	67.0	67.3	69.9	68.0	.0152	.5240
ST	66.0	63.0	63.5	64.7	66.5	.2034	.0010
GT	67.6	65.6	67.7	63.2	68.0	.5655	.0202

Standard Deviations

AFQT	10.2	10.5	11.8	10.8	12.5
CL	13.9	15.4	16.6	15.6	16.5
MM	11.4	11.4	11.3	12.2	11.7
SC	11.6	12.0	14.0	13.7	13.6
CO	12.4	9.9	10.2	11.6	11.9
FA	12.3	11.0	11.8	11.7	12.4
OF	11.5	11.2	11.8	11.7	12.4
GM	11.5	11.8	11.8	12.7	12.7
EL	12.4	12.7	11.8	12.9	13.5
ST	10.6	11.7	13.4	11.0	12.6
GT	12.8	12.9	13.2	13.1	13.5

Table 5.
INTER-ADMINISTRATION CORRELATIONS
OF THE TEN AREA COMPOSITES AND AFQT

Gen Maintenance (GM)

	Session 2	3	4	5
Ses. 1	.67	.83	.79	.82
2		.73	.77	.74
3			.81	.82
4				.85

Gen Tech (GT)

	Session 2	3	4	5
Ses. 1	.69	.69	.76	.72
2		.77	.76	.86
3			.77	.84
4				.79

Clerical (CL)

	Session 2	3	4	5
Ses. 1	.91	.88	.88	.88
2		.90	.91	.91
3			.94	.87
4				.91

Electronics (EL)

	Session 2	3	4	5
Ses. 1	.66	.80	.75	.85
2		.74	.81	.73
3			.82	.79
4				.84

Surv/Comm (SC)

	Session 2	3	4	5
Ses. 1	.90	.87	.85	.85
2		.89	.87	.87
3			.90	.87
4				.88

Motor Maintenance (MM)

	Session 2	3	4	5
Ses. 1	.79	.66	.72	.81
2		.62	.75	.77
3			.65	.65
4				.83

Field Artil (FA)

	Session 2	3	4	5
Ses. 1	.64	.75	.78	.76
2		.78	.74	.76
3			.87	.81
4				.85

Combat (CO)

	Session 2	3	4	5
Ses. 1	.67	.74	.82	.75
2		.76	.76	.73
3			.83	.81
4				.87

Skilled Tech (ST)

	Session 2	3	4	5
Ses. 1	.75	.79	.82	.83
2		.79	.80	.81
3			.83	.81
4				.86

Operators/Foods (OF)

	Session 2	3	4	5
Ses. 1	.80	.82	.87	.84
2		.79	.80	.78
3			.87	.85
4				.91

Table 5 (Cont.)

Armed Forces Qualification Test (AFQT)

	Session			
	2	3	4	5
Ses.				
1	.90	.91	.92	.92
2		.94	.92	.91
3			.93	.93
4				.92

APPENDIX A

A COMPARISON OF TEST/RETEST CORRELATIONS FOR THE PRESENT SAMPLE (N=57) EARLY (SESSIONS 1, 2) AND LATE (SESSIONS 4, 5) IN PRACTICE AND A REFERENCE SAMPLE TESTED TWICE

(Correlations Corrected for Restriction in Range)

Subtest	Sessions 1, 2	Army Sample[1]	Sessions 4, 5
GS	.8216	.7887	.8929
AR	.8681	.8649	.9258
WK	.8388	.8392	.8923
PC	.7266	.6261	.6837
NO	.8670	.7523	.9305
CS	.8360	.7115	.8806
AS	.8258	.7998	.9029
MK	.8842	.8656	.8952
MC	.7309	.7803	.8972
EL	.8535	.7351	.8409

---

1 Source: Friedman, Streicher, Wing & Grafton, 1982.